DOCUMENT RESUME

ED 202 897 TM 810 322

AUTHOR Popham, W. James
TITLE Measurement Concomitants of Competency-Oriented Instruction.
PUB DATE Apr 80
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Administrative Policy; Equated Scores; Graduation Requirements; *High Schools; *Minimum Competency Testing; *Program Implementation; Standards; *Testing Problems; Test Selection

ABSTRACT

Six measurement-related problems which face educators attempting to implement minimum competency-oriented instruction (MC-OI) are isolated. They are: (1) choosing appropriate competencies to measure; (2) acquiring appropriate tests; (3) setting standards for MC-OI; (4) selecting suitable sanctions; (5) providing equidifficult examination opportunities; and (6) linking competency tests to norm-referenced measures. Strategies are suggested for coping with each problem. When choosing competencies to measure, the media must be informed well in advance of the test-taking just what rationale underlies the selection of competencies. Teachers need tests chosen which provide cues for instructional decision-making. Whether standard-setting is to be considered a matter only for professional judgment, or a matter for community judgment, should be made clear at the outset to all those involved in the competency program. In an effort to avoid litigation from disgruntled parents of students who fail to pass a minimum competency test, some administrators are devising a system of including various kinds of endorsements on its diplomas. The Rasch model provides a solution to the test equating problem. Although it seems sensible to have a local criterion-referenced competency test satisfy federal requirements for norm-referenced tests, complexities and legal difficulties must first be overcome. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W. J. Popham

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED202897

MEASUREMENT CONCOMITANTS OF COMPETENCY-ORIENTED INSTRUCTION*

W. James Popham
University of California, Los Angeles
and
Instructional Objectives Exchange

In its most fundamental form, the scenario for a competency-oriented instructional endeavor resembles a typical three-act play. Act One: competencies are determined; Act Two: measures of the competencies are secured; Act Three: students are given instruction designed to promote mastery of those competencies. Assuming that these three acts are properly performed, one can visualize a successful educational drama followed by suitable applause and loads of curtain calls.

But anyone who has participated in the conduct of an actual competency-oriented instructional undertaking will surely agree that although there may be only three acts, each act contains a number of scenes. And, for many of us, those scenes embody broad new challenges. We're being called on to play roles we've not previously encountered. Competency-based educational programs are tossing obstacles at us that we're often ill-equiped to overcome.

In this analysis, I shall isolate a half-dozen measurement-related problems, puzzles, and perplexities which educators are currently discovering as they attempt to implement competency-oriented instructional endeavors. I have personally witnessed educators grappling with each of these problems in various parts of the U.S. during the past few years. In some cases the problems have been solved. In most cases, the grappling goes on. Some of these problems, being eminently predictable, have already been written about at some length. Others, much less predictable, are only now being recognized and have not, therefore, been seriously addressed. To keep this analysis manageable, I shall restrict my observations to minimum competency programs, the sort that we find abounding all over the country these days. In general, however, many of these same difficulties will be encountered in any competency-oriented instructional program, whether minimum, maximum, or medium. Let's

TM 810 322

*A symposium presentation at the annual meeting of the American Educational Research Association, April 7-11, 1980, Boston, Massachusetts.

initiate this undertaking with a consideration of three readily anticipatible perplexities.

Choosing Appropriate Competencies to Measure

In almost every minimum competency program ever established, educators initially attend to the isolation of the competencies which will constitute the focus of the program. It's such a logical first step. And, while most educators concede that the determination of competencies is not foolsplay, they generally have been rather satisfied with their efforts to choose suitable minimal competencies. Usually those competencies have consisted of basic skills in reading, writing, and arithmetic--often cast in ways that reflect life-role rather than merely academic skills. In many cases the competency selection process has involved rather elaborate input from a variety of concerned constituencies, such as teachers, citizens, and students. By and large, most educators have been pretty well satisfied with their competency-choosing efforts.

But in recent months, since many districts/states are now administering the tests and reporting results, we've begun to see public reactions to student performance on minimum competency tests. We find two distinctive public reactions to the competencies, neither one of them all that heartwarming.

In settings where formidable, i.e., high level, competencies have been selected, student performance on the tests to measure those competencies has often been rather dismal. The citizenry becomes irate at "poor pupil mastery of fundamental skills." The schools are chided by the media for their inability to transmit rudimentary skills in the three Rs. Educators get more bad press, another black eye, and a continued diminution of public confidence.

In settings where less taxing competencies have been chosen, the media often denigrate the rigor of the tests. Students, in post-examination newspaper interviews, dismiss the tests as "too easy" or "a joke." Thus, even though most students may pass the competency tests, the competencies are tagged with a "trivial" label. Educators get more bad press, another black eye, and a continued diminution of public confidence.

We have, it would seem, a classic Scylla-Charybdis dilemma. If educators select really taxing competencies, students' test scores are poor and the educators lose. If more pedestrian competencies are chosen, students' test scores are high and the educators lose. Where, oh where, is justice?

About the only solution strategy I have been able to come up with for this apparently no-win puzzle hinges on the use by local educators of an aggressive, proactive media campaign. The media must be informed well in advance of the test-taking just what rationale underlies the selection of competencies.

If, for example, high level competencies have been chosen, then citizens must be prepared for the likely event that at the outset of the competency program, many students may not do all that well. But citizens can be informed that the district/state has selected challenging competencies because it wishes to provide defensible aspirations for pupils. In time, as the educational system becomes more attuned to the promotion of those high level skills, a greater proportion of pupils will pass. In essence, the public should be prepared not to judge a book by its preface.

At the other extreme, if lower level competencies have been chosen, again the media should be given an advance message, but this time that message is quite different. Citizens should be informed that the minimum competencies are just that--minimum skills--and that they constitute only a necessary but not sufficient condition for the student's graduation (promotion, etc.). Citizens should be told that there are many more requisites (e.g., compulsory courses) that a student must satisfy prior to receipt of a diploma, but that mastery of minimum skills at least provides the public with a limited warranty that undeserving students are not rewarded.

Most educators who have participated in the competency-selection process recognize and understand the factors that inclined them to opt for higher or lower sorts of skills. Those sorts of insights must be shared with the public via the media. Given properly fashioned rhetoric and a rigorous "pre-test" communication effort with the media, many educators may just be able to hop the horns of this apparent dilemma.

Acquiring Appropriate Tests

In a race toward the rapid implementation of minimum competency programs (sometimes imposed by legal or policy means), many educators have chosen unwisely the tests to assess students' mastery of competencies. A fair number of these tests have been developed along traditional norm-referenced testing lines and bear few, if any, implications for instructional design. As a consequence, once the testing program gets underway, teachers find it difficult to create instructional programs designed to ameliorate learner deficiencies. The tests chosen were developed to discriminate among examinees, not provide teachers with cues for instructional decision-making.[1] What teachers need are tests that clearly stake out the territory, tests that describe what it is that is "fair game" for each competency. Many currently used minimum competency tests provide teachers with scant guidance regarding the eligible kinds of test items that can be used to assess a particular competency. The absence of tangible targets for teachers, of course, dooms the minimum competency program to a series of, at best, only incremental improvements. Such tests will only serve to produce an annual testimony to educational failure.

To overcome this difficulty, educators need not abandon the program's competencies (since they are often stated at a somewhat general level). Instead, they merely need jettison the unsuitable tests and replace them with tests that serve as a spur to improved instruction. Replacement of traditionally constructed tests with more modern, instructionally sensitive tests will not be all that simple. But, if the program is to prosper, such substitutions are imperative.

Standard-Setting Anxieties

In the last few of years, numerous authors have dealt with the difficulties of establishing defensible standards for minimum competency programs. For example, the entire issue of the Winter 1978 Journal of Educational Measurement was devoted to a consideration of issues surrounding the standard-setting operation. Even with all this attention, however, we're still a long way from satisfactory resolution of the many procedural and conceptual difficulties associated with the setting of standards for minimum competency programs.

---

[1]For a description of tests that offer cues for instructional design, see Popham, W. James, "Educational Measurement for the Improvement of Instruction," Phi Delta Kappan, in press.

But educators still are obliged to set standards for minimum competency programs, by hook, crook, or any other ook-ending procedure. In the process of doing so, we find some educators stumbling on a particularly troublesome problem which might be characterized as local ownership versus "local ownership."

Perhaps the issue can best be illuminated via an actual illustration. A large school district in the process of implementing a minimum competency program for high school graduation had carried out the early developmental phases of the program in a remarkably open fashion. Teachers, students, and citizens were heavily involved in the selection of the minimum competencies and in a number of decisions about operational aspects of the program. It had been anticipated, and publicly announced, that when it came time for standard-setting, those concerned constituencies would once more be called in to offer advice regarding appropriate cut-off scores.

But as the time for standard-setting drew near, the administrators in the district grew increasingly uneasy about allowing laypersons to become involved in that phase of the operation. As a result, when the district's governing board finally received a recommendation regarding cut-off scores, it stemmed from the closed-door deliberations of a relatively small group of high level administrators--deliberations not based on the anticipated widespread community involvement.

The lesson to be learned from this example, it would seem, is that the designers of a minimum competency system must, in advance, mentally play out the entire drama to see if they can comfortably live with community input at the various points where it might reasonably be appropriate. If standard-setting, because of its obviously important implications for the program's image of success, is considered a matter only for professional judgment, then this should be made clear at the outset to all those involved in the competency program.

If widespread local ownership of the minimum competency program is really being fostered, it should be genuine. If there are stages in the competency program's evolution where only insiders' decisions count, then outsiders ought to know that so they can decide what degree of participation they wish to invest in the remaining phases of the enterprise.

Let's turn now to three less predictable measurement-related problems which educators are now encountering as they become involved with their minimum competency maelstroms.

Selecting Suitable Sanctions

What happens when a student fails a minimum competency test repeatedly? What negative sanctions are visited on that student? In the early days of the minimum competency testing movement this issue seemed readily resolvable. Students who passed the test would receive a high school diploma. Students who didn't pass the test would be given no diploma, receiving at the most some sort of "certificate of attendance" or the like.

Now, however, perhaps in an effort to avoid litigation from disgruntled parents of non-diploma students, some educators are devising a variety of second place prizes for those failing minimum competency tests. For instance, one school district is weighing the merits of various kinds of endorsements on its diplomas, so that while students judged competent receive fully endorsed and suitably embellished diplomas, those judged incompetent receive diplomas without that endorsement. Thus, every student who completes routine academic requirements receives a high school diploma, it's just that some diplomas carry more prestige than others. One can visualize districts creating a multilevel battery of diplomas ranging from "super" to "sluggish."

Although it is too early to discern the effects of these differential diploma schemes, it is clear that inventive educators will doubtlessly devise a score of alternatives to the standard diploma. Hopefully, educators experimenting with novel sorts of consolation prizes, or novel ways of boosting the prestige of the regular diploma, will share their experiences regarding the pros and cons of these approaches--particularly with respect to the manner in which they are received by pupils, parents, teachers, and the community at large.

Equidifficult Examination Opportunities

In the early days of minimum competency testing we saw educators racing to complete legislatively set requirements of their programs. To select competencies, get a test, and install remedial programs took more time then was available. The intensity of effort was often fevered. But, by and large, such tasks were accomplished.

In year two of the program, however, another batch of students had to be examined, and those failing the first year's test had to be given a second assessment opportunity. Given probable breeches of security if the first year's test was simply re-administered, a second test form was needed. And therein lies a substantial problem for the implementers of minimum competency testing programs, namely, how to come up with several equidifficult test forms.

Although there's a fair amount of literature about how to ascertain the equivalent-forms reliability of tests, there's really not all that much information on the step-by-step procedures one should follow in whomping up mirror-image test forms. Educators have been discoverying that the problem is far more difficult than they had imagined. How can equidifficult test forms be created by busy educators? How, indeed, can equidifficult test forms be created so that the school system is safe from legal assault by parents of youngsters who received a "too difficult" form of a supposedly equidifficult test?

Well, one way of generating equidifficult test forms would hinge on marathon testing. In this approach we could generate huge item pools, corral a large number of students, then have all the students complete all the items so that for each item we would obtain a p value (the proportion of examinees answering an item correctly). Different forms of the test would then be constituted which were equal both in content as well as difficulty (as reflected by mean p values).

The problems with this approach are several. First off, it is very costly to generate huge item pools at one time. If we could produce those additional items over an extended period of time, the costs would typically be spread out and thus more tolerable. Second, fatigue factors might be operative after students had completed a hundred or so test items, thereby rendering their responses atypical. Finally, because the students would realize that their marathon test-taking session was not the "real thing," and only of use in test-equating, their unmotivated responses to items would often be less authentic than would be the case if the stakes (e.g., a high school diploma) were real. The marathon approach to test-equating is, obviously, fraught with problems.

Another scheme for equating tests is to administer one form of a test, then subsequently generate enough new items so that they could be

field tested on a different population of examinees in order to secure test items of equivalent difficulty (p values).

For example, suppose a 60-item test of mathematics achievement had been given as part of a minimum competency testing program to all of the 2,000 eleventh-graders in a school district, with a resulting mean p value of .71.

Let's say we cranked out 100 new items and administered them to a different sample of 300 next year's eleventh-grade students so that we could put together a new form of the test whose content coverage matched that of the first test and whose mean p value was also .71. Doesn't that sound pretty good so far?

But there is a flaw. If the 300 students used in the new tryout of the 100 items are not identical in ability to the original 2,000 eleventh-graders, then misleading p values will be yielded. Thus, even if the second test has an average p value of .71, it may be substantially different than the first version of the test. For instance, suppose the 300 youngsters were substantially less bright than the original 2,000. Then a second test with a mean p value of .71 (based on weaker students' performances) would be a markedly easier test than the earlier test.

Similarly, let's say that there had recently been highly effective mathematics instruction in the district so that the 300 eleventh-graders (although equal in intelligence to the 2,000 students) really performed well on the test. Again, even though the average p values on both tests are .71, the second test will turn out to be even tougher because the sample students used in the item selection are much more knowledgeable than the original students who didn't have the benefit of the improved instruction. A similar situation would also arise if the second test had been tried out on 300 students who were much brighter than the original 2,000.

The dilemma, then, is trying to develop equivalent test forms so that the artificiality or the exhaustion associated with a single group of students' taking many test items fails to yield p values of comparable meaning. On the other hand, when we use different samples of examinees in an effort to secure comparable p values, differences in the ability or training of those samples can lead to non-comparable p values.

Rasch as the Remedy

The Rasch model, developed some years ago by the Danish mathematician Georg Rasch, provides us with a way of overcoming these difficulties and therefore developing different test forms which, if not equivalent in difficulty, can permit us to make comparable judgments as to how examinees would have fared as if the exams had been identical in difficulty. Although space and time prohibit an elaboration of the particulars by which Rasch approaches permit educators to solve the multiple test forms problem, a number of state and local school systems are sufficiently convinced of the Rasch model's potency in this regard that they are committed to its use. The Detroit Public Schools and the South Carolina Department of Education, for example, have adopted a Rasch solution to the test equating problem.

Linking Competency Tests to Norm-Referenced Measures

To close out my half-dozen problems, I wish to raise an issue that only recently seems to be surfacing. Since many states and districts are now administering minimum competency tests on a routine basis to many of their pupils, they find themselves wishing to use results on those tests to satisfy certain federal dollar-dispensing requirements. Unfortunately, the federal requirements often call for norm-referenced tests, but many competency tests are criterion-referenced (often having been produced locally).

In order to cope with this situation, some school systems have begun to explore the possibility of somehow hooking up their tests with a nationally standardized norm-referenced test so that the local criterion-referenced test could be used in lieu of the norm-referenced measure.

While it is conceptually simple to set forth the rules for creating equivalency tables to suit this mission, there are some tricky complexities associated with the generation of new forms of the competency test (and field-testing items, some of which may not end up on the test), then correlating examinees' results on both tests. In addition, there are legal difficulties associated with the use of copyrighted test norms for such purposes.

In other words, although it seems to be a sensible thing to have a local criterion-referenced competency test do double-duty, thereby reducing measurement costs and the intrusiveness of testing on students'

instructional programs, there are many wrinkles to be smoothed out before we accomplish this modern measurement coup.

In review, I have given a once-over-lightly treatment to six measurement-related problem areas which have emerged from the installation of minimum competency testing programs. Clearly, there are many more such problems which warrant our serious analysis. But, for openers, these half-dozen puzzles might be considered, well, a minimum.

58:WJP